Homework Assignment #1
Due at midnight Sunday, 10/2

Part-1

Programming ID-3 like Decision Tree Learner.
Implementation source code files submitted. (main file: *dt-learn.py*)

Part-2

Plot learning curves that characterize the predictive accuracy of learned trees as a function of the training set size.

a) Heart Domain:

Plotting source code included in *dt-learn.py* (Method: *plot_accuracy_vs_training_set_size()* with appropriate training and test data files *heart_train.arff* and *heart_test.arff*).
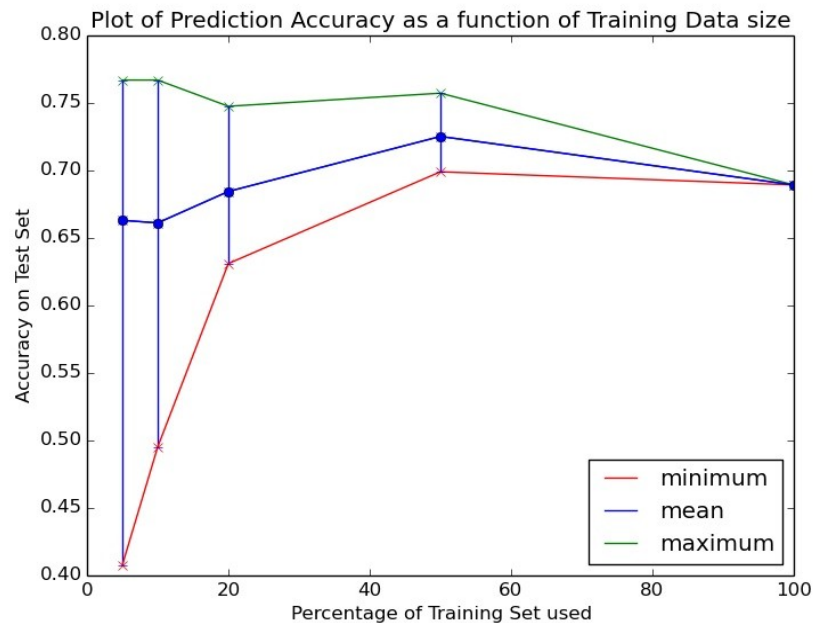
Percentage of Training Set Used	Minimum Predictive Accuracy	Mean Predictive Accuracy	Maximum Predictive Accuracy
5	0.40776699	0. 6631068	0.76699029
10	0.49514563	0.66116505	0.76699029
20	0.63106796	0.68446602	0.74757282
50	0.69902913	0.72524272	0.75728155
100	0.68932039	0.68932039	0.68932039

b) Diabetes Domain:

Plotting source code included in *dt-learn.py* (Method: *plot_accuracy_vs_training_set_size()* with appropriate training and test data files *diabetes_train.arff* and *diabetes_test.arff*).
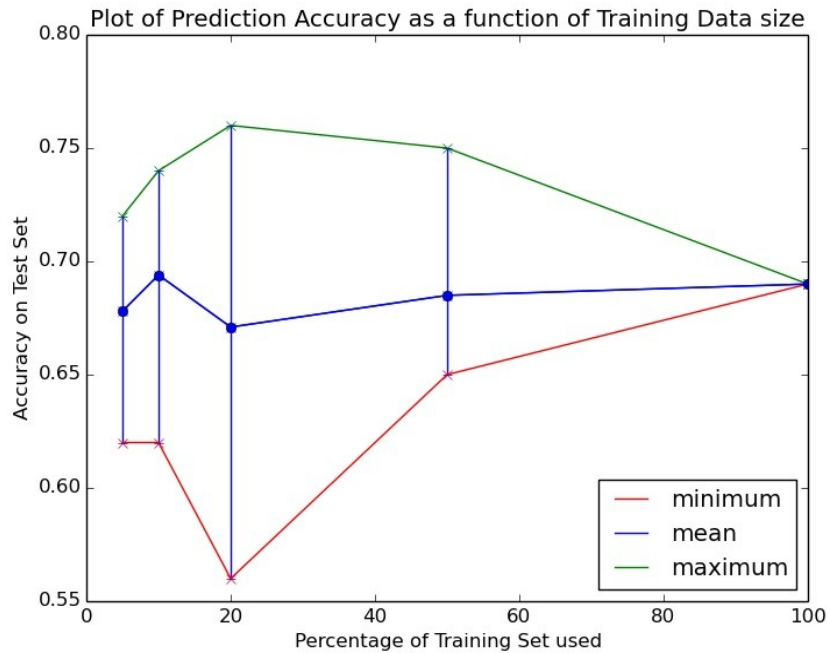

Percentage of Training Set Used	Minimum Predictive Accuracy	Mean Predictive Accuracy	Maximum Predictive Accuracy
5	0.62	0.678	0.72
10	0.62	0.694	0.74
20	0.56	0.671	0.76
50	0.65	0.685	0.75
100	0.69	0.690	0.69

Part-3

Investigate how predictive accuracy varies as a function of tree size. Plot curves showing how test-set accuracy varies with the value *m* used in the stopping criteria.

a) Heart Domain:

Plotting source code included in *dt-learn.py* (Method: *plot_accuracy_vs_tree_size()* with appropriate training and test data files *heart_train.arff* and *heart_test.arff*)
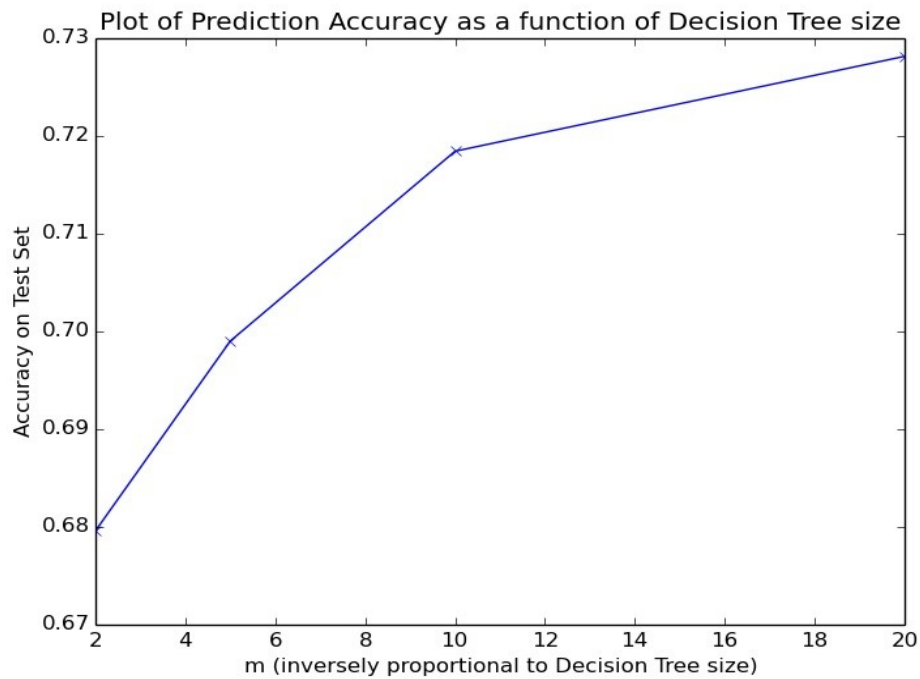
m (stopping criteria)	Predictive Accuracy
2	0.6796116504854369
5	0.6990291262135923
10	0.7184466019417476
20	0.7281553398058253

b) Diabetes Domain:

Plotting source code included in *dt-learn.py* (Method: *plot_accuracy_vs_tree_size()* with appropriate training and test data files *diabetes_train.arff* and *diabetes_test.arff*).
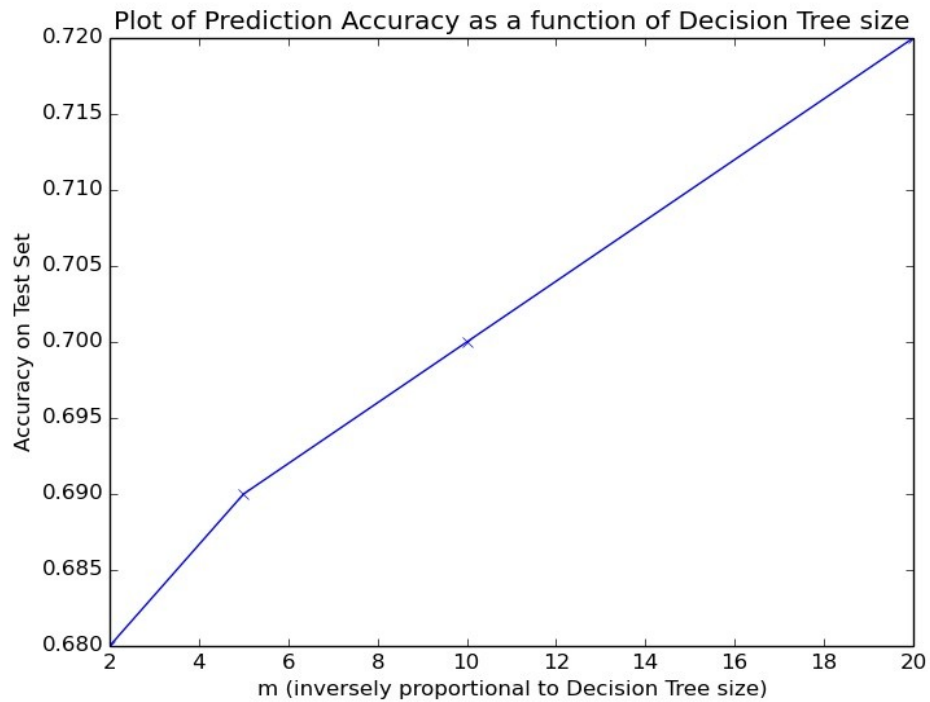
m (stopping criteria)	Predictive Accuracy
2	0.68
5	0.69
10	0.70
20	0.72